DOI:
10.1053/j.gastro.2018.09.038

Document Version
Peer reviewed version

Accepted Manuscript

Information on Genetic Variants Does Not Increase Identification of Individuals at Risk of Esophageal Adenocarcinoma Compared to Clinical Risk Factors

A.T. Kunzmann, M. Canadas Garre, A.P. Thrift, Ú.C. McMenamin, B.T. Johnston, C.R. Cardwell, L.A. Anderson, A.D. Spence, J. Lagergren, S.-H. Xie, L.J. Smyth, A.J. McKnight, H.G. Coleman

PII: S0016-5085(18)35032-7
DOI: 10.1053/j.gastro.2018.09.038
Reference: YGAST 62148

To appear in: *Gastroenterology*
Accepted Date: 13 September 2018

Please cite this article as: Kunzmann A, Canadas Garre M, Thrift A, McMenamin Ú, Johnston B, Cardwell C, Anderson L, Spence A, Lagergren J, Xie S-H, Smyth L, McKnight A, Coleman H, Information on Genetic Variants Does Not Increase Identification of Individuals at Risk of Esophageal Adenocarcinoma Compared to Clinical Risk Factors, *Gastroenterology* (2018), doi: https://doi.org/10.1053/j.gastro.2018.09.038.

Title: Information on Genetic Variants Does Not Increase Identification of Individuals at Risk of Esophageal Adenocarcinoma Compared to Clinical Risk Factors

Authors:

AT Kunzmann[1], M Canadas Garre[2], AP Thrift[3], ÚC McMenamin[1], BT Johnston[4] CR Cardwell[1], LA Anderson[1], AD Spence[1], J Lagergren[5,6], S-H Xie[5], LJ Smyth[2], AJ McKnight[2] & HG Coleman[1,7]

Authors' affiliations:

[1] Cancer Epidemiology Research Group, Centre for Public Health, Queen's University Belfast, Belfast, Northern Ireland, United Kingdom

[2] Epidemiology and Public Health Research Group, Centre for Public Health, Queen's University Belfast, Belfast, Northern Ireland, United Kingdom

[3] Section of Epidemiology and Population Sciences, Department of Medicine, Baylor College of Medicine, Houston, Texas

[4] Royal Victoria Hospital, Belfast Health & Social Care Trust, Belfast, N. Ireland, United Kingdom

[5] Upper Gastrointestinal Surgery, Department of Molecular medicine and Surgery, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden

[6] School of Cancer and Pharmaceutical Sciences, King's College London, London, United Kingdom

[7] Centre for Cancer Research and Cell Biology, Queen's University Belfast, Belfast, Northern Ireland, United Kingdom

Corresponding author: Dr Andrew Kunzmann. Mailing address: Centre for Public Health, Queen's University Belfast, Institute of Clinical Sciences-B, Royal Victoria Hospital Site, Grosvenor Rd, Belfast, Northern Ireland, BT12 6BJ. E-mail: a.kunzmann@qub.ac.uk.

Grant support: The project was kindly funded by **Ochre charity (**Registered charity number: SC032343). MCG is funded by a Science Foundation Ireland-Department for the Economy (SFI-DfE) Investigator Program Partnership Award (15/IA/3152).

Disclosures: The authors disclose no potential conflicts of interest.

Author contributions:

ATK: study concept and design; acquisition of data; analysis and interpretation of data; drafting of the manuscript; critical revision of the manuscript for important intellectual content; statistical analysis; obtained funding.

MCG: acquisition of data; analysis and interpretation of data; statistical analysis; critical revision of the manuscript for important intellectual content.

APT: critical revision of the manuscript for important intellectual content.

ÚCM: acquisition of data; critical revision of the manuscript for important intellectual content.

BTJ: critical revision of the manuscript for important intellectual content.

CRC: acquisition of data; analysis and interpretation of data; critical revision of the manuscript for important intellectual content.

LAA: critical revision of the manuscript for important intellectual content.

ADS: critical revision of the manuscript for important intellectual content.

JL: critical revision of the manuscript for important intellectual content.

SX: critical revision of the manuscript for important intellectual content.

LJS: critical revision of the manuscript for important intellectual content.

AJM: critical revision of the manuscript for important intellectual content.

HGC: study concept and design; acquisition of data; critical revision of the manuscript for important intellectual content; study supervision.

Abstract

We previously developed a tool that identified individuals who later developed esophageal adenocarcinoma (EAC; based on age, sex, body mass index, smoking status, and prior esophageal conditions) with an area under the curve of 0.80. In this study, we collected data from 329,463 individuals in the UK Biobank cohort who were tested for genetic susceptibility to EAC (a polygenic risk score based on 18 recognized genetic variants). We found that after inclusion of this genetic information, the area under the curve for identification of individuals who developed EAC remained at 0.80. Testing for genetic variants associated with EAC therefore seems unlikely to improve identification of individuals at risk of EAC.

Keywords: esophagus; cancer; early detection; stratification; mutation

Novel screening and risk-stratification methods are needed to improve early detection of esophageal adenocarcinoma as the vast majority of patients are currently diagnosed at a late stage when survival is poor[1] and population-wide endoscopy screening is unlikely to be cost-effective[2]. Established clinical risk factors may be useful in identifying individuals at a higher risk of esophageal adenocarcinoma[4-6]. Our findings within the UK Biobank indicated that a model including age, sex, body mass index (BMI), smoking status and prior esophageal conditions had an area under receiver operating characteristic curve (AUROC) for predicting esophageal adenocarcinoma within 5 years of 0.80 (95% CI 0.76-0.82)[3]. Whilst these results suggested that clinical risk factors may be useful as an inexpensive and non-invasive risk-stratification tool, additional robust, minimally-invasive follow-up risk-stratification methods will likely be required to identify individuals at sufficiently high risk to warrant endoscopy screening. Conversely, those at lowest risk could avoid unnecessary and invasive procedures.

Germline mutations associated with esophageal adenocarcinoma risk can be readily assessed using DNA derived from minimally invasively biological samples, thus their evaluation for the purpose of risk stratification in a routine setting is appealing. The clinical utility of assessing single nucleotide polymorphisms (SNPs) for esophageal adenocarcinoma risk-prediction has been investigated in a previous case-control study consortium[6], but remains to be investigated in large-scale prospective studies.

In this study, we analyzed prospective data from the UK Biobank to assess whether the addition of genetic factors to an established clinical risk prediction score enhances the ability to identify individuals at high risk of esophageal adenocarcinoma development within 5 years. A secondary aim was to validate the associations between SNPs and esophageal adenocarcinoma development identified in large genome-wide association studies (GWAS) [7-8].

Clinical factors were assessed using a touchscreen questionnaire, and BMI measurements and blood tests were taken at baseline[9] . The clinical risk-factors (age ,sex, BMI, smoking status and reflux

diagnosis/symptoms) used for risk-prediction modelling were selected from a previous investigation within the UK Biobank[3]. Germline mutations were genotyped using the Affymetrix UK BiLEVE and UK Biobank Axiom arrays (Santa Clara, CA, USA), followed by imputation methods to >90 million variants. Previous GWAS identified 24 genetic variants associated with esophageal adenocarcinoma[7-9]. Two SNPs were unavailable for analysis (rs9918259 and rs75783973), two SNPs were excluded due to linkage disequilibrium (Supplementary Table 1) and one SNP (rs9257809) was excluded as it was not in Hardy-Weinberg equilibrium (P<0.01). Polygenic risk scores were calculated by summing the positive risk allele counts for 18 genetic variants weighted by their odds ratios from previous GWAS [8-9], and dividing the total by 18[6].

Diagnostic accuracy was quantified for the combined genetic and clinical risk factor models using the AUROC curve with 95% confidence interval (CI). A points-based model, created from our previous investigation of established clinical risk factors[3], was used to assign points to tertiles of the polygenic risk score. Sensitivity, specificity, Youden's index (sensitivity + specificity -1) and net reclassification index[10] were assessed for individuals above each points based cut-off threshold when using the established risk factor model alone and when combined with the genetic factors. Further details are outlined in the Supplementary Methods.

Of 502,640 participants in UK Biobank, the following exclusions were applied: 117,891 were aged under 50 years, 30,665 had a history of cancer (or cancer within 6 months of baseline), 8,515 had missing clinical or genetic data, two participants withdrew consent, 772 had insufficient genetic data quality and 15,152 reported a non-White ethnicity. A total of 329,643 (65.6%) were eligible for inclusion, among whom 214 individuals were diagnosed with esophageal adenocarcinoma within 5 years. Individuals diagnosed with esophageal adenocarcinoma were more likely to be older, male, smoke (current or former), have a higher BMI, have an existing esophageal condition, and have a higher polygenic risk score than non-cases (Supplementary Table 2). Mean follow-up time was 4.8 (standard deviation 0.6) years.

The strength of associations for each individual SNP were broadly in line with expectations [7-9], though modest and not statistically significant, in either crude or multivariable analyses (Supplementary Table 3). The strongest adjusted associations with esophageal adenocarcinoma within 5 years were for SNPs at, or near, the genes *CFTR* (rs17451754; OR 1.21, 95% CI 0.80-1.83) and *BARX1* (rs11789015; OR 1.20, 95% CI 0.96-1.50).

The association between polygenic risk score and risk of esophageal adenocarcinoma within 5 years was modest and not statistically significant (Adjusted OR $_{\text{middle versus bottom tertile}}$ 1.09, 95% CI 0.77-1.55; Adjusted OR $_{\text{top versus bottom tertile}}$ 1.38, 95% CI 0.99-1.92).

The AUROC for the clinical risk factor model was 0.80 (95% CI 0.76-0.82), as previously published[3], and was unchanged when polygenic risk score categories were added to the model (0.80, 95% CI 0.77-0.83) (Figure 1) and in a secondary analyses adding the 18 SNPs to the model (0.81, 95% CI 0.78-0.83).

A points-based model, used points created from a previous investigation of established clinical risk factors[3], i.e. age (55-60 years: 1.5; 60-65 years: 2.5; 65+ years: 3.5), sex (males: 4), smoking status (former: 2; current: 3.5), BMI (>25-30: 1; 30-<35: 1.5; 35+: 2.5), history of esophageal conditions or treatment (1.5). Points for polygenic risk score categories were assigned (middle category=0 points, top category=1 point) by dividing their coefficient when added to the clinical risk factor model by the smallest coefficient in the previous model (0.40 for BMI of 25-<30 kg/m^2) and rounding to the nearest 0.5 to allow easier to interpret cut-offs. When comparing the combined clinical and polygenic risk score model to the original clinical points based model, changes in net reclassification index[10] and Youden's index[11] were modest at any of the cut-off points, as modest improvements in sensitivity were offset by modest reductions in specificity (Table 1).

The results from this large UK cohort study with prospective follow-up data suggest that SNPs previously implicated in esophageal adenocarcinoma susceptibility do not aid risk-prediction alone, or in conjunction with known clinical risk factors. The lack of predictive ability occurred regardless of

method used to derive the polygenic risk score and seems unlikely to be improved by additional statistical power as AUROC confidence intervals were narrow. These results validate findings from a previous case-control study[6], despite the weaker association between polygenic risk scores and oesophageal adenocarcinoma in the current study. Similar lack of improvements were apparent when stratified by certain demographic features (Supplementary Table 4). Future genetic analyses of families with a history of esophageal adenocarcinoma may be required to identify a polygenic risk score that accounts for a larger proportion of the heritability of esophageal adenocarcinoma than the SNPs identified using GWAS studies to date.

Nevertheless, the results provide modest support for a potential role of some genes in esophageal adenocarcinoma development. In particular, genetic variants on or near the CFTR gene (rs17451754) related to cystic fibrosis, which displays phenotypic overlap with reflux[12], and the *BARX1* gene (rs11789015) related to differentiation of esophageal epithelia[13].

To our knowledge, this is the first prospective assessment of genetic susceptibility combined with known clinical risk factors in esophageal adenocarcinoma risk prediction. Potential limitations include reduced statistical power to detect significant associations for low frequency SNPs; lack of information for two SNPs previously identified in GWAS or family history of esophageal adenocarcinoma and; criticism of the UK Biobank's generalizability[14].

Thus, testing for currently recognized germline mutations is unlikely to improve stratification of individuals at risk of esophageal adenocarcinoma at a population level. Future studies should examine other novel screening methods, biomarkers, or epigenetic studies to achieve earlier detection of this tumor.

References

1. Coleman HG et al. Gastroenterology 2017; 154: 390–405
2. Lagergren J et al. Lancet 2017; 390: 2383–96
3. Kunzmann AT et al. Clinical gastroenterology and hepatology 2018
4. Thrift AP et al. Clinical Gastroenterology and Hepatology 2013; 11: 138–144.e2
5. Xie S-H et al. The American Journal of Gastroenterology 2018; 113: 829–35
6. Dong J et al. Gastroenterology 2018; 154: 1273–1281.e3
7. **Gharahkhani P, Fitzgerald RC, Vaughan TL, Palles C, Gockel I, Tomlinson I** et al. The Lancet Oncology 2016; 17: 1363–738. Levine DM et al. Nature Genetics 2013; 45: 1487–93
9. Allen N et al. Health Policy and Technology 2012; 1: 123–26
10. Pencina MJ et al. Statistics in Medicine 2008; 27: 157–72
11. Youden WJ. Cancer 1950; 3: 32–35
12. Meltzer SJ. The Lancet Oncology 2016; 17: 1336–37
13. Woo J et al. PLoS ONE 2011; 6: e22493
14. Fry A et al. American Journal of Epidemiology 2017; 186: 1026–34
15. Purcell S et al. The American Journal of Human Genetics 2007; 81: 559–75

Author names in bold designate shared co-first authorship

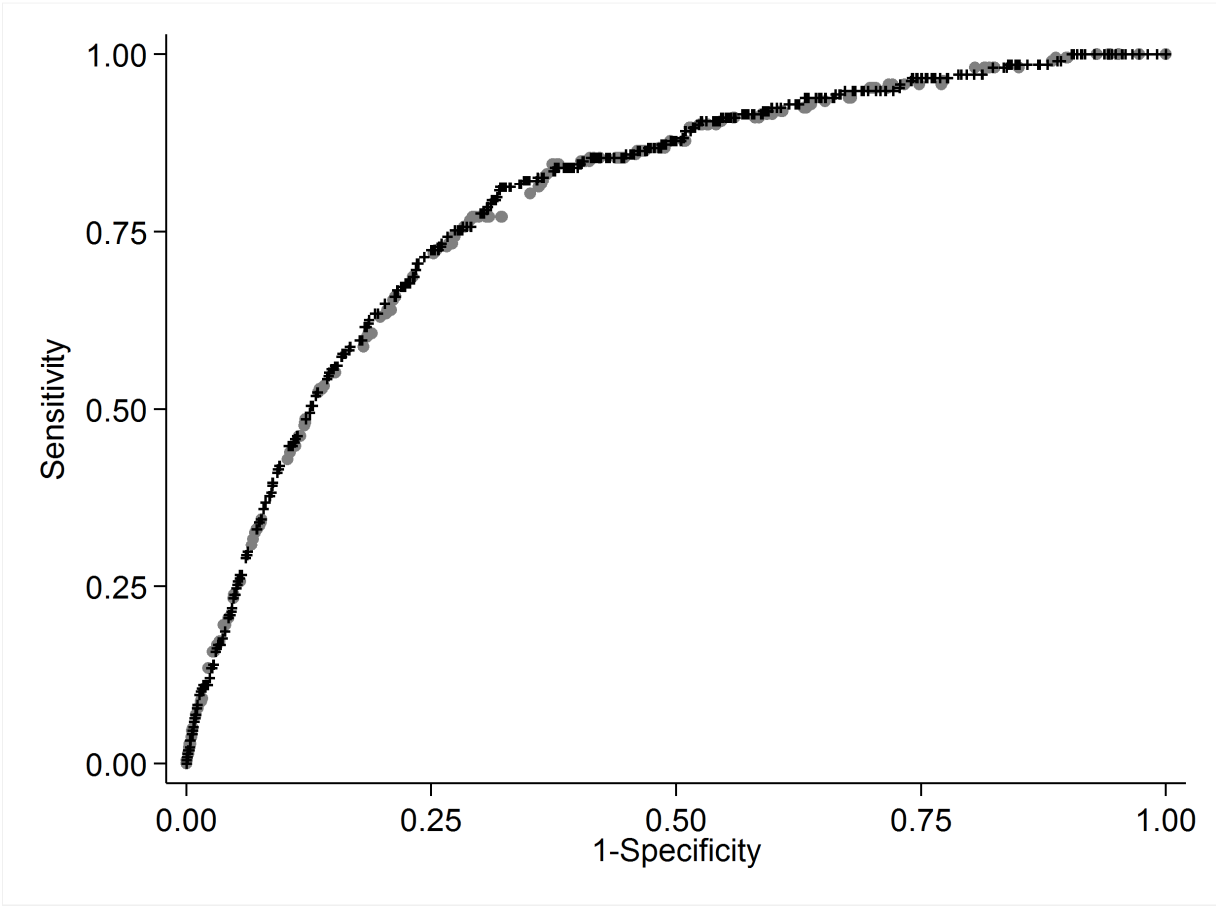
Figure legend

Figure 1. Receiver operating characteristic curve for the clinical model alone (•) and when combined with polygenic risk score categories (+) for predicting risk of esophageal adenocarcinoma within 5 years.

Table 1. Performance statistics of the combined (clinical and polygenic risk scores) and clinical only points-based esophageal adenocarcinoma risk-prediction models at different points based cut-offs

Points cut-off		Patients deemed high-risk (%)	Sensitivity	Specificity	Youden's index	NRI[1]
7+	Combined	146,042 (44.3%)	86.4%	55.7%	0.42	-0.04
	Original	134,550 (40.8%)	85.0%	59.2%	0.44	-
8+	Combined	110,358 (33.5%)	81.3%	66.5%	0.48	0.01
	Original	98,203 (29.8%)	77.1%	70.2%	0.47	-
9+	Combined	78,215 (23.7%)	69.2%	76.3%	0.46	0.05
	Original	68,383 (20.8%)	63.6%	79.2%	0.43	-
10+	Combined	52,439 (15.9%)	57.9%	84.1%	0.42	0.07
	Original	44,350 (13.5%)	51.9%	86.5%	0.38	-

[1] Net Reclassification index: positive values indicate that a larger proportion of individuals with events were moved up to the high-risk group than individuals without events when changing from the original to combined model.



Supplementary methods

Study design

This cohort study used prospective data from the UK Biobank, which recruited 502,640 men and women aged 40-69 years from one of 22 centers located across England, Scotland, and Wales between 2006 and 2010[9]. Approximately 9.2 million individuals registered with the National Health Service living within a 25-mile (~40km) radius of one of the 22 centers were invited to participate. The response rate was 5.5% (n=503,325)[14]. Included in the present study were individuals reporting a white ethnicity who had not withdrawn consent, aged ≥50 years (as upper gastrointestinal cancers are rare aged <50), without a history of cancer (excluding non-melanoma skin cancer) at or before baseline or within 6 months following baseline (to exclude diagnostic delays), and with complete information on relevant risk factors and SNPs.

The UK Biobank was approved by the North West Multi-Centre Research Ethics Committee, and all participants provided written informed consent.

Assessment and classification of clinical and genetic risk-factors

Participants were asked to complete electronic touchscreen questionnaires at baseline, which enquired about a wide range of potential risk factors for chronic diseases (including age, sex, smoking status and prior medical conditions); have anthropometric measurements taken and; provide a blood test for genetic analysis at baseline.

UK Biobank samples were genotyped using the Affymetrix UK Biobank Axiom array (Santa Clara, CA, USA) or the similar Affymetrix UK BiLEVE Axiom array (the former is an update of latter, and the two arrays share 95% content). Quality control, phasing and imputation through the IMPUTE3 program were carried out centrally.

Hardy-Weinberg equilibrium and pairwise haplotype frequencies were estimated, and Lewontin's D prime (D') and the linkage disequilibrium coefficient (r^2) were calculated.

Previous genome-wide association studies (GWAS) identified 24 genetic variants associated with esophageal adenocarcinoma[7-9]. Two of these genetic variants (rs9918259 and rs75783973) identified in previous GWAS[7-8] were not imputed so were unavailable for this analysis. Two further SNPs were excluded due to linkage disequilibrium with other SNPs in the analysis (rs76014404, R^2=0.96 and rs199620551, R^2=0.77, Supplementary Table 1) and one SNP (rs9257809) was excluded as it was not in Hardy-Weinberg equilibrium (P<0.01). Polygenic risk scores were calculated by summing the positive risk allele counts for the 18 variants weighted by their odds ratios for associations with esophageal adenocarcinoma from previous GWAS[7-9], and dividing the total by 18. The polygenic risk score categories were created using tertiles.

Outcome assessment

The UK Biobank is regularly linked to UK cancer registry data from the Health and Social Care Information Centre (in England and Wales), the Scottish Cancer Registry (in Scotland) and death records from the UK Office of National Statistics (ONS). Cancer data were provided up until 30th September 2014. Newly diagnosed cancers were classified by site according to International Classification of Diseases, 10th version (ICD/10) and histology (ICD-O morphology codes). Primary esophageal adenocarcinoma (ICD/10 C15, with ICD-O 8140–8573) diagnosed between 6 months (due to potential diagnostic delays) and 5 years from baseline was the main outcome of interest. The 5-year follow-up may have reduced statistical power but may offer a more clinically useful time-point to initiate screening.

Statistical analysis

Logistic regression was used to assess the association between established clinical risk factors alone, and in combination with polygenic risk score categories, and risk of esophageal adenocarcinoma within 5 years. Diagnostic accuracy was quantified for the clinical factor and combined genetic and factor models using the AUROC curve with 95% confidence interval (CI).

A points-based model, created from a previous investigation of established clinical risk factors [3], was used to assign points to tertiles of the polygenic risk score by dividing by the smallest coefficient in the previous model (0.40 for BMI of 25-<30kg/m^2) and rounding to the nearest 0.5 to allow ease of calculation without a computer and easier to interpret cut-offs. Sensitivity, specificity, Youden's index (sensitivity + specificity -1) and net reclassification index[10] were assessed for individuals above each points based cut-off threshold when using the established risk factor model alone and when combined with the genetic factors.

Logistic regression was used to assess the association between individual SNPs and risk of esophageal adenocarcinoma within 5 years, with and without inclusion of established clinical risk factors or other SNPs. Diagnostic accuracy was quantified for the model including all 18 SNPs and clinical risk factors model using the AUROC curve with 95% confidence interval (CI).

Stratified analyses by age, sex, BMI, smoking history and prior esophageal conditions were conducted to assess whether the change in diagnostic accuracy when adding the polygenic risk scores to the established clinical factors was improved in certain high risk groups.

SNPs were extracted and linkage disequilibrium, Hardy-Weinberg equilibrium and minor allele frequencies were calculated using PLINK [15]. All other analyses were conducted using Stata/SE statistical software (version 14.1, College Station, TX, USA).

Supplementary Table 1. Linkage disequilibrium

Chr	GENE	SNP1	SNP2	R^2	D'
2	*GDF7& SATB2*	rs7255	rs13397172	<0.01	<0.01
2	*SATB2 & GDF7*	rs13397172	rs3072	<0.01	<0.01
2	*GDF7*	rs7255	rs3072	0.63	0.97
3	*FOXP1 & HTR3C*	rs2687202	rs9823696	<0.01	0.02
6	*KHDRBS2-MTRNR2L9 & MHC region*	rs62423175	rs9257809	<0.01	0.01
6	*MHC region & KHDRBS2*	rs9257809	6:62391538_TAAACA_T	<0.01	<0.01
6	*KHDRBS2*	rs62423175	6:62391538_TAAACA_T	<0.01	0.94
7	*CFTR & ASZ1*	rs17451754	rs2188554	0.48	0.85
8	*LINC00208-BLK & MSRA*	rs10108511	rs17749155	0.11	0.71
9	*MSRA*	rs7852462	rs11789015	<0.01	<0.01
15	*ALDH1A2*	rs2464469	15:58267416_GACAT_G	0.46	0.89
19	*CRTC1*	19:18804294_TG_T	rs10419226	0.96	0.99
19	*CRTC1*	rs10419226	rs10423674	0.39	0.96
19	*CRTC1*	19:18804294_TG_T	rs10423674	0.42	0.98

Abbreviations: Chr: Chromosome; SNP: Single Nucleotide Polymorphism; R^2: linkage disequilibrium coefficient; D': Lewontin's D prime.

Supplementary Table 2. Characteristics of study population who did or did not develop esophageal adenocarcinoma within 5 years

		No esophageal adenocarcinoma Number (%)	Esophageal adenocarcinoma Number (%)
	Total	329,429	214
Age (years)			
	50-<55	65,442 (19.9)	14 (6.5)
	55-<60	78,488 (23.8)	39 (18.2)
	60-<65	105,024 (31.9)	74 (34.6)
	65+	80,475 (24.4)	87 (40.7)
Sex			
	Female	176,640 (53.6)	34 (15.9)
	Male	152,789 (46.4)	180 (84.1)
Smoking status			
	Never	172,236 (52.3)	55 (25.7)
	Former	125,886 (38.2)	114 (53.3)
	Current	31,307 (9.5)	45 (21.0)
Body mass index			
	<25	103,963 (31.6)	35 (16.4)
	25-<30	143,518 (43.6)	97 (45.3)
	30-<35	59,421 (18.0)	55 (25.7)
	35+	22,527 (6.8)	27 (12.6)
Esophageal condition[1]			
	No	284,404 (86.3)	156 (72.9)
	Yes	45,025 (13.7)	58 (27.1)
Polygenic risk score tertiles			
	Low	109,820 (33.3)	61 (28.5)
	Medium	109,814 (33.3)	67 (31.3)
	High	109,795 (33.3)	86 (40.2)

[1] Esophageal conditions included self-reported history of gastroesophageal reflux disease, Barrett's esophagus, hiatus hernia or esophageal stricture and/or; esophageal fundoplication or hiatus hernia surgery and/or; anti-reflux medication use (none or any)

Supplementary Table 3. The association between individual SNPs (previously associated with esophageal adenocarcinoma or Barrett's esophagus) and risk of esophageal adenocarcinoma within 5 years in the UK Biobank.

SNP	Position:	*Gene*	Minor Allele Frequency	Risk allele[1]	Published OR/risk allele[2]	Crude OR (95% CI)	Multivariable[3] OR (95% CI)
rs7255	2: 20878820	*GDF7-LDAH*	44% (T)	T	1.17	1.18 (0.97-1.42)	1.17 (0.85-1.61)
rs13397172 [4]	2: 200045039	*SATB2*	44% (T)	C	1.13	1.04 (0.86-1.26)	1.05 (0.86-1.27)
rs3072	2: 20878406	*GDF7*	36% (C)	C	1.14	1.15 (0.94-1.39)	1.00 (0.73-1.39)
rs2687202	3: 70929983	*FOXP1*	29% (T)	T	1.13	1.08 (0.88-1.32)	1.07 (0.87-1.32)
rs9823696	3: 183783353	*HTR3C*	38% (A)	A	1.17	0.97 (0.80-1.18)	0.97 (0.79-1.18)
rs62423175	6: 62195368	*KHDRBS2-MTRNR2L9*	17% (A)	A	1.23	0.96 (0.74-1.25)	0.96 (0.74-1.25)
rs17451754	7: 117256712	*CFTR*	13% (A)	G	1.25	1.31 (0.96-1.78)	1.21 (0.80-1.83)
rs2188554	7: 117040117	*ASZ1*	18% (G)	A	1.23	1.23 (0.94-1.60)	1.10 (0.77-1.57)
rs10108511	8: 11435516	*LINC00208-BLK*	46% (T)	T	1.12	1.14 (0.94-1.38)	1.15 (0.94-1.40)
rs17749155	8: 10068073	*MSRA*	16% (A)	A	1.14	1.03 (0.79-1.33)	0.95 (0.73-1.26)
rs7852462	9: 100310501	*TMOD1*	39% (T)	C	1.08	1.15 (0.94-1.41)	1.15 (0.94-1.40)
rs11789015	9: 96716028	*BARX1*	28% (G)	A	1.20	1.22 (0.97-1.52)	1.20 (0.96-1.50)
rs1247942	12: 114673723	*LOC105369996-TBX5*	41% (C)	G	1.11	1.11 (0.91-1.35)	1.11 (0.91-1.35)
rs2464469	15: 58362025	*ALDH1A2*	41% (G)	G	1.11	1.08 (0.89-1.31)	1.15 (0.88-1.50)
rs66725070	15: 58267416	*ALDH1A2*	46% (G)	GACAT	1.15	1.00 (0.83-1.21)	0.90 (0.69-1.18)
rs1979654	16: 86396835	*LOC732275-FOXF1*	41% (G)	G	1.11	0.84 (0.69-1.03)	0.84 (0.69-1.02)
rs10419226	19: 18803172	*CRTC1*	46% (T)	T	1.18	0.93 (0.77-1.12)	0.85 (0.67-1.08)
rs10423674	19: 18817903	*CRTC1*	33% (A)	C	1.19	1.04 (0.85-1.28)	1.15 (0.89-1.48)

[1] Allele associated with increased esophageal adenocarcinoma risk

[2] Odds ratios for association with esophageal adenocarcinoma in previous GWAS and used in creation of polygenic risk scores in the current study.

[3] Adjusted for age, sex, body mass index, smoking status and gastroesophageal reflux diagnosis/symptoms and other SNPs.

[4] Formerly referred to as *rs139606545* [6]

Abbreviations: Chr: Chromosome; CI: Confidence interval; OR: Odds ratio; SNP: Single Nucleotide Polymorphism.

Supplementary Table 4. The diagnostic accuracy when using the original model alone and combined with the polygenic risk score in stratified analyses.

	Original model AUC (95% CI)	Original model + PRS AUC (95% CI)	p-value for difference
BMI			
Normal BMI	0.83 (0.76-0.89)	0.84 (0.78-0.90)	0.18
Overweight or obese	0.78 (0.75-0.82)	0.79 (0.76-0.82)	0.37
Smoking			
Never	0.78 (0.72-0.84)	0.78 (0.72-0.84)	0.26
Ever	0.76 (0.73-0.79)	0.77 (0.73-0.80)	0.38
Prior oesophageal condition			
No	0.80 (0.76-0.83)	0.80 (0.77-0.84)	0.14
Yes	0.75 (0.69-0.81)	0.76 (0.70-0.81)	0.44
Sex			
Men	0.73 (0.69-0.76)	0.73 (0.70-0.76)	0.15
Women	0.71 (0.62-0.79)	0.71 (0.63-0.79)	0.82
Age			
Under 60	0.83 (0.78-0.88)	0.84 (0.79-0.88)	0.16
Over 60	0.76 (0.72-0.80)	0.76 (0.73-0.80)	0.43